NORMAL DEGREE DAYS ABOVE ANY BASE BY THE UNIVERSAL TRUNCATION COEFFICIENT

H. C. S. THOM
Environmental Data Service, Environmental Science Services Administration, Washington, D.C.

ABSTRACT

Equations are developed for obtaining mean monthly degree days above any base from mean monthly temperature and standard deviation of monthly average temperature. By the use of data for all months for twelve widely scattered stations and four bases it is shown that the truncation coefficient for degree days below any base with proper modification of the argument also applies for degree days above any base. This is also proved analytically, which leads to some further aspects of the universality of the truncation coefficient. Two formulas for the coefficient are also developed.

1. INTRODUCTION

The study reported here is the third phase in the development of the general climatological analysis of degree days [1], [2], [3]. Reference [2] established the rational relationship between mean monthly degree days below 65° F. and temperature and gave a table of the truncation coefficient l. Reference [3] showed that the rational relationship also applied to mean monthly degree days below any base using the same table of l. Later it was noted [4] that a slightly modified relationship using the same table of l also gave mean monthly degree days above any base. Thus the table of the truncation coefficient proved to be universal, applying to mean monthly degree days below or above any base. The evidence for the final steps in establishing universality has never been given although the method has been used extensively in this country and Canada [5], [6]. It is the purpose of this paper to give this evidence as well as an analytical form for the universal truncation coefficient useful in computer applications.

Degree days above particular bases, although not as yet used as extensively as degree days below a base, are of growing importance in horticulture and in air conditioning requirements and power consumption estimations. Horticulturists use bases ordinarily between 40° and 50° F. in systems for estimating growth progress and harvest dates. The literature on this application is extensive, of which reference [5] is a good example. Application to air conditioning requirement and power consumption has been much less extensive and even less has been published. The key paper is the one by Marston [6]. Indications are that there will be an increase in the use of degree days in this area.

2. DISTRIBUTION FUNCTION AND EXPECTED VALUE

It was shown previously that the degree day distribution describes a mixed population of degree day values equal to zero and greater than zero. This arises from the definition of the degree day; a particular value of which is the number of degrees of temperature above (or below) a fixed base temperature. Thus the temperature distribution truncated at the base temperature transformed to degree days, the continuous part of the distribution, and the truncated portion, the probability of zero degree days, form the mixed distribution of degree days. For degree days above a given base b the transformation from temperature to degree days is

$$D=t-b; \quad (D\geqq 0). \tag{1}$$

where t is ordinarily the average temperature for a day.

The truncated probability density function for temperature may be expressed by

$$f(t|b\leq t)=\frac{f(t)}{\int_b^\infty f(t)dt}=\frac{f(t)}{1-F(b)}, \tag{2}$$

where F is the distribution function of t, and the probability density function has the value given by (2) on the interval $b\leqq t<\infty$, and zero elsewhere. If the transformation (1) is applied to equation (2) in the usual fashion, the result is the probability density function of degree days

$$g(D|D\geq 0)=\frac{f(b+D)}{1-F(b)}\frac{dt}{dD}=\frac{f(b+D)}{1-F(b)}. \tag{3}$$

Integrating this over the open interval $0<D<\infty$ gives the distribution function of degree days greater than zero

$$G(D|D>0)=\frac{1}{1-F(b)}\int_{0+}^{D}f(b+D)dD=\frac{F(b+D)-F(b)}{1-F(b)}. \tag{4}$$

Multiplying by $1-F(b)$ and adding $F(b)$ gives the desired distribution function on the closed interval $0\leqq D<\infty$, i.e., including the zero values of degree days.

As in the first work on the rational relationship between degree days and temperature the analysis is performed on a hypothetical middle day of a month. The average temperature on this day is assumed to have a normal distribution whose mean and standard deviation are such that when the conversion is made to degree days, and the result multiplied by the number of days in the month, the result is the mean degree days for the month [2]. The normal probability density function is designated by φ and the distribution function by Φ.

Let the standardized variable of temperature be $z=[t-E(t)]/\sigma$ where $E(t)$ and σ are the population mean and standard deviation and the truncation point is z_0. Then the probability density function for the truncated normal distribution according to equation (2) may be expressed as

$$\varphi(z|z_0\leq z)=\frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}/[1-\Phi(z_0)]. \tag{5}$$

Hence the mean of this distribution is given by

$$E(z|z_0<z)=\frac{1}{\sqrt{2\pi}[1-\Phi(z_0)]}\int_{z_0}^{\infty} ze^{-z^2/2}dz. \tag{6}$$

To evaluate the integral it is only necessary to make the substitutions $u=-z^2/2$ and $dz=-du/z$ whence

$$-\int e^u du=-e^u=-e^{-z^2/2}.$$

Substitution in (6) and evaluation of the integral between z_0 and ∞ yields the reciprocal Mill's ratio

$$E(z|z_0<z)=\frac{e^{-z_0^2/2}}{\sqrt{2\pi}[1-\Phi(z_0)]}=\frac{\varphi(z_0)}{1-\Phi(z_0)}=\lambda_*(z_0) \tag{7}$$

where the inferior star indicates truncation on the left of the distribution. To return to the variable t it is only necessary to take the mean of $t=z+E(t)$ over the truncated distribution giving

$$E(t|b<t)=\sigma E(z|z_0<z)+E(t) \tag{8}$$

which on substituting (7) yields

$$E(t|b<t)=\sigma\lambda_*(z_0)+E(t). \tag{9}$$

The mean number of degree days greater than zero is found by taking the expected value of $D=t-b$, giving $E(D|D<0)=E(t|b<t)-b$ which on substitution of equation (9) gives

$$E(D|D>0)=\sigma\lambda_*(z_0)+E(t)-b. \tag{10}$$

The mean of degree days for the mixed population of zero and non-zero degree days is the weighted mean of these components or

$$E(D|D\geq 0)=\Phi(z_0)\cdot 0+[1-\Phi(z_0)]E(D|D>0).$$

Substituting from (10) gives

$$E(D|D\geq 0)=[1-\Phi(z_0)][\sigma\lambda_*(z_0)+E(t)-b] \tag{11}$$

which is the theoretical relationship between mean temperature and degree days for a middle day with mean $E(t)$ and standard deviation σ. Unfortunately, estimates of σ and therefore of z_0 are not available; so, as for degree days below a base [2], an approximation of σ must be employed.

3. GENERAL DEGREE DAY FORMULA

With σ unavailable it is necessary to make some adjustment to equation (11) which makes computation possible. The most suitable procedure was found to be to follow the method used for degree days below a base, i.e., to solve as much as possible for $\lambda_*(z_0)$ and associated functions of z_0 which are not known.

Rearranging the terms in equation (11) and writing $E(D)$ for $E(D|D\geq 0)$ yields

$$\lambda_*(z_0)-\frac{E(D)}{\sigma}\left[\frac{\Phi(z_0)}{1-\Phi(z_0)}\right]=\frac{E(D)-[E(t)-b]}{\sigma}. \tag{12}$$

As with degree days below a base, the left hand side is set equal to a new truncation coefficient Λ_* after a modification of the right hand side to take care of the fact that no direct estimate is available for σ. Let σ_m be the standard deviation of monthly average temperature and $\bar{\rho}$ the mean correlation between all possible pairs of N days of a month; then

$$\sigma=\sqrt{N}\sigma_m/[1+(N-1)\bar{\rho}]^{1/2}. \tag{13}$$

The factor $[1+(N-1)\bar{\rho}]^{1/2}$ is unknown because $\bar{\rho}$ is unknown, but call it k nevertheless so (13) becomes

$$\sigma=\sqrt{N}\sigma_m/k. \tag{14}$$

In order to standardize the argument on which Λ_* is dependent, $E(t)$, b, and $\sqrt{N}\sigma_m$ are combined into a single term to make the standardized truncation point

$$-x_0=(\bar{t}-b)/(\sqrt{N}\sigma_m). \tag{15}$$

This was $-h$ of the previous paper [2]. Now since k is unknown, replace σ on the right of (12) by $\sqrt{N}\sigma_m$ and let the factor k divide the term on the left. Finally replace the left hand term by $\Lambda_*(x_0)$ so that

$$\Lambda_*(x_0)=\frac{E(D)-[E(t)-b]}{\sqrt{N}\sigma_m}. \tag{16}$$

This is the population value of the truncation coefficient for degree days above a base b. Solving for $E(D)$ and multiplying by N to get the monthly mean degree days above b gives

$$NE(D)=N[\Lambda_*(x_0)\sqrt{N}\sigma_m+E(t)-b]. \tag{17}$$

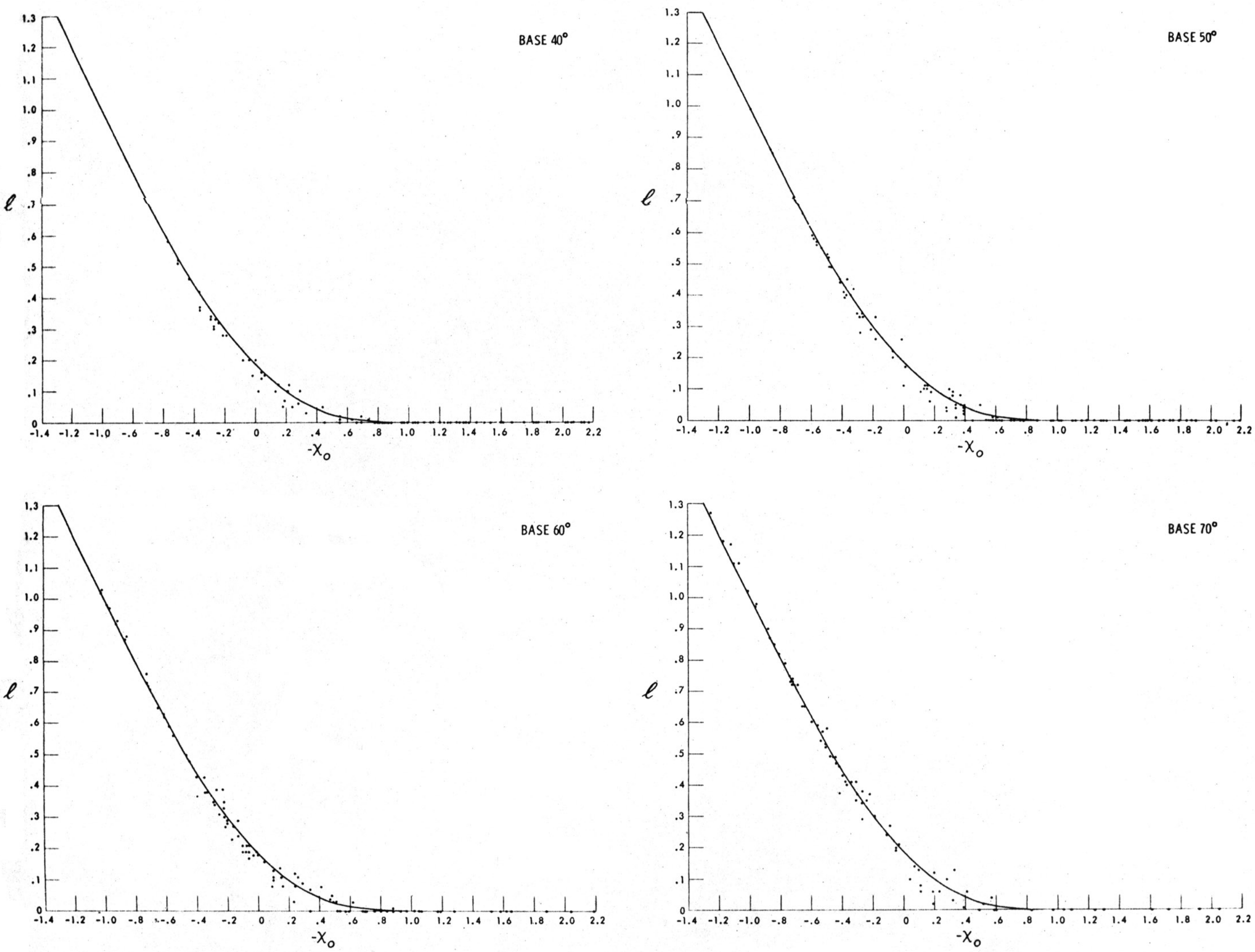

Figure 1.—l vs. x_0 data for various bases. The l-curve for degree days below 65° F. is superimposed on data for above various bases to show universality of the l-curve.

The sample estimates of these two equations are

$$l_*(x_0)=\frac{\overline{D}-(\bar{t}-b)}{\sqrt{N}s_m} \tag{18}$$

and

$$N\overline{D}=N[l_*(x_0)\sqrt{N}s_m+\bar{t}-b]. \tag{19}$$

It was now immediately conjectured that l plotted against $-x_0$ or $-h$ would produce an l curve independent of the base and identical with the l curve previously established for degree days below any base [3].

To show the universality of the truncation coefficient l_*, values were computed using equation (17) on degree day means above four bases for all months having degree days at 12 widely scattered stations. These are shown plotted against $-x_0$ in figure 1. The l-curve established for degree days below 65° F. [2] and found to hold for degree days below any base [3] was then superimposed on the data for the four bases in figure 1. The fit is equally as good as found for degree days below a base. This completes the empirical demonstration of the universality of the truncation coefficient for mean monthly degree days above or below any base. The demonstration will be made analytically in the next section.

It is clear from the above that if the previous h is set equal to x_0, representing the truncation coefficient for degrees below a base as $l^*(x_0)$ then empirically at least

$$l_*(x_0)=l^*(-x_0). \tag{20}$$

Substituting in (19) gives

$$N\overline{D}=N[l^*(-x_0)\sqrt{N}s_m+\bar{t}-b]. \tag{21}$$

Writing $\overline{D}^{+}$ for degree days above a base and $\overline{D}^{-}$ for degree days below a base generalizes (21) to

$$N\overline{D}^{\pm}=N[l^{*}(\mp x_0)\sqrt{N}s_m\pm(\bar{t}-b)] \tag{22}$$

which covers all situations.

4. UNIVERSAL TRUNCATION COEFFICIENT

In this section Λ^* and Λ_* are always a function of x_0 and λ^* and λ_* is still a function of z_0.

For degree days above a base there are two expressions for $E(D)$ given by equation (11) and an inversion of (16). This gives

$$\Lambda_*\sqrt{N}\sigma_m+(E(t)-b)=(1-\Phi)[\lambda_*\sigma+(E(t)-b)]. \tag{23}$$

Dividing by $\sqrt{N}\sigma_m$, substituting the value of k from equation (14), and rearranging terms yield finally

$$\Lambda_*=(1-\Phi)(\lambda^*/k)+\Phi x_0. \tag{24}$$

For degree days below a base [2] there are analogous expressions for $E(D)$ which when set equal give

$$\Lambda^*\sqrt{N}\sigma_m-(E(t)-b)=\Phi[\lambda^*\sigma-(E(t)-b)]. \tag{25}$$

Again dividing by $\sqrt{N}\sigma_m$, etc., as above, yields finally

$$\Lambda^*=\Phi(\lambda^*/k)-(1-\Phi)x_0. \tag{26}$$

If equations (23) and (25) are divided by σ instead of $\sqrt{N}\sigma_m$ using, of course, $z_0=[b-E(t)]/\sigma$ and manipulations similar to those above, there results

$$\Lambda_*=[(1-\Phi)\lambda_*+\Phi z_0]/k \tag{27}$$

and

$$\Lambda^*=[\Phi\lambda^*-(1-\Phi)z_0]/k. \tag{28}$$

Setting the value of Λ_* from (24) equal to that of (28) gives

$$z_0=kx_0 \tag{29}$$

which is also clear from the definitions of x_0 and z_0.

The basic equations (24) and (26) may be transformed by recalling that $\lambda^*=\varphi/\Phi$ and $\lambda_*=\varphi/(t-\Phi)$. Substituting yields

$$\Lambda^*=\varphi/k-(1-\Phi)x_0 \tag{30}$$

and

$$\Lambda_*=\varphi/k+\Phi x_0. \tag{31}$$

Since $\varphi(-x_0)=\varphi(x_0)$, φ is not affected by a change of sign of its argument. It is noted from equation (29) that k must also be an even function of x_0, thus the first terms of (30) and (31) are not affected by a change in sign of x_0. Returning for the moment to explicit expressions for $\Lambda^*(x_0)$ and $\Phi(x_0)$ and substituting $-x_0$ for x_0 in (31) yield

$$\Lambda^*(-x_0)=\varphi/k-[1-\Phi(-x_0)](-x_0). \tag{32}$$

Recalling that $\Phi(-x_0)=1-\Phi(x_0)$ and making this substitution in (32) give

$$\Lambda^*(-x_0)=\varphi/k+\Phi x_0. \tag{33}$$

But this is identical with (31); hence

$$\Lambda^*(-x_0)=\Lambda_*(x_0) \tag{34}$$

which demonstrates the universality of the truncation coefficient. Starting from equation (31) and following similar operations give

$$\Lambda_*(-x_0)=\Lambda^*(x_0). \tag{35}$$

There are a number of other symmetrical relations which are interesting: Subtracting equation (30) from equation (33) yields the relation

$$\Lambda^*(-x_0)-\Lambda^*(x_0)=x_0. \tag{36}$$

Likewise, following similar operations or simply substituting (34) and (35) in (36) gives

$$\Lambda_*(x_0)-\Lambda_*(-x_0)=x_0. \tag{37}$$

These relations indicate the fundamental properties of the truncation function which assist in establishing its analytical form.

5. ANALYTICAL FORMS OF THE TRUNCATION CURVE

The truncation curve is not a very simple function as can be seen from the previous development. Since for practical applications it need only be known to two significant figures, it seemed reasonable to fit a curve to the l-table given in [2] taking into account the symmetry properties of the previous section.

None of the functional forms related to Mill's ratio proved to be of much help. Finally, it was found that the sum of two exponentials gave a very satisfactory result. Fitting to the original l-table gave the following pair of equations:

$$l^*(x_0)=0.34e^{-4.7x_0}-0.15e^{-7.8x_0} \tag{38}$$

and by (36)

$$l^*(-x_0)=l^*(x_0)+x_0.$$

These equations smoothed the l-table slightly. Departures from the unsmoothed table are not greater than 0.01.

It appeared to be of interest to relate the truncation function to Mill's ratio. It is necessary now to use x_0 as the independent variable for all functions. Solving equation (30) for k yields

$$k(x_0)=\frac{\varphi(x_0)}{\Lambda^*(x_0)+[1-\Phi(x_0)]x_0} \tag{39}$$

Recalling that $\varphi(x_0)$ is an even function and substituting $-x_0$ for x_0 give

$$k(-x_0)=\frac{\varphi(x_0)}{\Lambda^*(x_0)+[1-\Phi(x_0)]x_0} \tag{40}$$

Mill's ratio is defined for this purpose as

$$R(x_0)=[1-\Phi(x_0)]/\varphi(x_0). \tag{41}$$

Solving for $[1-\Phi(x_0)]$ and substituting in (39) give

$$k(x_0)=\frac{\varphi(x_0)}{\Lambda^*(x_0)+\varphi(x_0)R(x_0)x_0}. \tag{42}$$

Since it is required to fit $k(x_0)$ for both positive and negative values of x_0, it is necessary to have a formula for $k(-x_0)$. Solving (41) for $\Phi(x_0)$ and substituting in (40) yield

$$k(-x_0)=\frac{\varphi(x_0)}{\Lambda^*(-x_0)+\varphi(x_0)R(x_0)-1}. \tag{43}$$

A series of values of $l^*(x_0)$ and $l^*(-x_0)$, $x_0=h$, for each tenth between 1.00 and -2.00 was obtained from the l-table of [2]. Values of $\varphi(x_0)$ and $R(x_0)$ were found in tables II and III in reference [7]. When these values are substituted in equations (42) and (43) a series of $k(x_0)$ values is obtained. Note that the positive value of x_0 is always used in $R(x_0)$.

Examination of equations (13) and (14) suggests that it might be more interesting to determine the equation for k^2 instead of k since $k^2=1+(N-1)\bar{\rho}$. The series of values obtained from equations (42) and (43) were therefore squared before being fitted as a function of x_0.

After a new series was formed by subtracting one from each k^2 a functional form $k^2-1=y=a\cos^n\theta$ was intuitively suggested. If $\tan\theta=x_0$, $\theta=\tan^{-1}x_0$, for $x_0=-2.0$, -1.0, and 0, $\theta=-1.11$, -0.785, and 0 radians, hence $\cos\theta=0.4474$, 0.7071, and 1. Since k^2-1 is about 3.410 at $\theta=0.785$ radians, $1.326=3.410(0.7071)^n$ and $n=2.73$. (This will incidentally be very close to the final value.) An approximation to the equation is then $y=3.410\cos^{2.73}\theta$. On substituting $\theta=\tan^{-1}x_0$ and by simple trigonometry we find $y=3.410[1+x_0^2]^{-1.36}$. The general form of the equation is then

$$k^2-1=a[1+x_0^2]^{-m}. \tag{44}$$

The logarithmic form of this was fitted by least squares giving finally

$$k^2(x_0)=1+3.44(1+x_0^2)^{-1.35}. \tag{45}$$

The fit of this to the l-table was very good, for the correlation between the logarithms was $r^2=0.9897$, leaving only about 1 percent of the variance unexplained by equation (45). With the k-function in analytical form a second method of computing l^* using Mill's ratio is available.

REFERENCES

1. H. C. S. Thom, "Seasonal Degree-Day Statistics for the United States," *Monthly Weather Review*, vol. 80, No. 9, Sept. 1952, pp. 143–149.
2. H. C. S. Thom, "The Rational Relationship between Heating Degree Days and Temperature," *Monthly Weather Review*, vol. 82, No. 1, Jan. 1954, pp. 1–6.
3. H. C. S. Thom, "Normal Degree Days Below Any Base," *Monthly Weather Review*, vol. 82, No. 5, May 1954, pp. 111–115.
4. H. C. S. Thom, "Standard Deviation of Monthly Average Temperature," *U.S. National Atlas*, 1955, pp. 1–123.
5. R. M. Holmes and G. W. Robertson, "Heat Units and Crop Growth," *Publication 1042*, Canada Department of Agriculture, 1959, 31 pp.
6. A. D. Marston, "Degree Days for Summer Air Conditioning," Kansas City Power and Light Company, Report, 1937.
7. K. Pearson (editor), *Tables for Statisticians and Biometricians*, Part II, Cambridge University Press, London, 1931, pp. 2–10 and 11–15.

[*Received April 27, 1966*]